**By Team Statistics:**

|  |  |  |
| --- | --- | --- |
| Offense Score | New England | 29.94 |
| First Downs | New England | 21.58 |
| 3rd Down Percentage | Indianapolis Colts | 44.33 |
| Rush Yards | Denver Broncos | 132.12 |
| Yards Per Rush | Minnesota Vikings | 4.67 |
| Pass Yards | New Orleans Saints | 269.41 |
| Pass Completion Percent | Indianapolis Colts | 64.02 |
| Yards Per Pass | Pittsburgh Steelers | 12.15 |
| Interceptions | New England Patriots | 0.75 |
| Fumbles | Indianapolis Colts | 0.58 |
| Win Proportion | New England Patriots | 0.73 |

**Top 5 Teams (Excluding Win Prop) (In Order):**

1. New England Patriots
2. Denver Broncos
3. Indianapolis Colts
4. Green Bay Packers
5. Pittsburgh Steelers

**Bottom 5 Teams:**

5) Detroit Lions

4) Buffalo Bills

3) Chicago Bears

2) Arizona Cardinals

1) Cleveland Browns

**By Season Statistics:**

* General increase over time for good vars (so, cpo)
* General decrease over seasons for bad vars (pio, fo)
* 3rd Down Percentage stayed relatively the same
* YPP went down

**SCRIPT:**

Hello everyone, my name is Vidusha Rao. I am a student at the Institute for Advanced Analytics, and I will be talking about NFL data and statistics that will not help you win Fantasy Football.

The first thing that needs to be said is that from 2000 to 2013, the New England Patriots had a 73% Win Percentage. That is nuts. I am by no means a Patriots fan and I am definitely not a Tom Brady fan, but I have to respect the fact that they know how to win. I feel gross hyping up the Patriots but as an aspiring data scientist I have represent my findings honestly.

On to the data. The NFL data set I worked with, after combining all the data sets, had 7156 observations and 31 variables and spanned 14 seasons. I then ordered the data by seasons and filtered out variables that were not relevant to statistics related to Offense (leaving me with 13 variables). Even though Defense wins championships, I filtered out the Defense-related data as there would be a high degree of alignment with the other teams’ Offense data. I then grouped the data by season, then by team, and then completed logistic regression analysis on the variables.

For the season grouping, I looked for trends over time for different variables. Some variables showed a positive trend, some showed a negative trend, and some showed trends flatter than the footballs that were used during Deflate-Gate.

3rd Down Conversion Percentages stayed relatively consistent over the 14-year time frame. It seems that clutch situations have not been improved upon during this time period.

Average Fumbles per Game and Average Interceptions per Game showed a visible decline over the 14 seasons.

Average Score per Game and Average Pass Completion Percentage per Game both showed increased over the time frame. My best guesses for explaining these patterns are:

1. Offenses became more structured and organized, which helped to minimize the errors made and improve upon the amount of successful plays
2. Defenses have been deteriorating in quality over time and cannot keep up with Offense-related skills

These guesses come from a guy who spends more time coding about the NFL than actually watching it, so who knows what the actual reasons are.

Next, I will show you the wonderful world of bar graphs and logistic regression.

The first graph shows the Average Score per Game for all the NFL teams (Red teams belong to the AFC, blue to the NFC) along with the logistic regression plot. Higher scores have higher predicted probabilities of winning, but having more points does not guarantee a win. If the opposing team scores more points, you lose the game, regardless of how much you scored.

Here is an example of a negative relationship with fumbles. The predicted probabilities of winning the game decrease as the number of fumbles per game increase.

This slide shows which team had the best average for each variable. As you can see, New England hogs the leaderboard as they have the top stats for 4 out of the 11 variables.

Using the averages for each variable grouped by NFL team, I created a ranking system to rank the team’s based on their stats. The top 5 teams for the 2000-2013 seasons were (Read top 5 list) and the Bottom 5 teams were (read bottom 5 list).

Some steps I could have taken if I had more time would be to forecast the statistics to predict how teams would perform in the future and compare those forecasts to the actual outcomes.

You now have some fun facts about football in the 2000s that you can tell your friends. I hope you enjoyed this presentation. Thanks for listening!